

Transfer of Training on the Vertical Motion Simulator

Peter M.T. Zaal*
*San Jose State University
NASA Ames Research Center
Moffett Field, CA*

Jeffery A. Schroeder†
*Federal Aviation Administration
Moffett Field, CA*

William W. Chung‡
*Science Applications International
Corporation
NASA Ames Research Center
Moffett Field, CA*

This paper describes a quasi-transfer-of-training study in the NASA Ames Vertical Motion Simulator (VMS). Sixty-one general aviation pilots trained on four challenging commercial transport tasks under one of four different motion conditions: no motion, small hexapod, large hexapod, and VMS motion. Then, every pilot repeated the tasks in a check with VMS motion to determine if training with different motion conditions had an effect on task performance. New objective motion criteria guided the selection of the motion parameters for the small and large hexapod conditions. Considering results that were statistically significant, or marginally significant, the motion condition used in training affected 1) longitudinal and lateral touchdown position; 2) the number of secondary stall warnings in a stall recovery; 3) pilot ratings of motion utility and maximum load factor obtained in an overbanked upset recovery; and 4) pilot ratings of motion utility and pedal input reaction time in the engine-out-on-takeoff task. Since the training motion conditions revealed statistical differences on objective measures in all the tasks performed in the VMS motion check, with some in the direction not predicted, trainers should be cautious not to oversimplify the effects of platform motion. Evidence suggests that the new objective motion criteria may offer valid standardization benefits, as increases in the training motion fidelity, as predicted by the two conditions covered by the criteria, resulted in expected trends in pilot ratings and objective performance measures after transfer.

I. Introduction

This paper addresses two questions. First, “does objective evidence exist for the value of training with platform motion?” To date, the answer has been negative,^{1,2} and some airlines have requested training credit in situations without platform motion that regulators have historically required platform motion. The second question is “are recently proposed objective motion criteria effective in determining the value of motion for training?” These criteria³ may add to the current simulator motion evaluation practice, which is predominantly subjective.^{4,5} Validated objective criteria could promote standardization and minimize possible inappropriate training.⁶

Considerable research underlies both questions, yet gaps remain that this paper addresses. For the first question, previous transfer-of-training studies have not validated their findings in a transport category aircraft. Previous studies used either general aviation or military aircraft, or simulators in a quasi-transfer paradigm. This study does not use a transport category aircraft either, but applies the quasi-transfer paradigm on a transport category aircraft simulator with an order-of-magnitude more physical motion than in a typical training simulator. For the second question, the objective motion criteria are new, so this study is the first test of their effectiveness prior to their possible incorporation into regulation.

The paper unfolds as follows. First, a brief overview of relevant transfer-of-training research is given. A description of the four relatively challenging tasks, which try to increase the odds of revealing motion effects, follows the overview. Then, the paper describes the experimental design, which used the world’s largest flight simulator configured to test conditions ranging from no motion, through two hexapod configurations, to its full capabilities. The two hexapod configurations reasonably span the new objective motion criteria. After presentation of the results, a discussion follows. The paper ends with conclusions and references.

*Research Associate, Human Systems Integration Division, NASA Ames Research Center, Moffett Field, CA, 94035; peter.m.t.zaal@nasa.gov. Member.

†Chief Scientific and Technical Advisor for Flight Simulation Systems, Federal Aviation Administration, Moffett Field, CA, 94035; jeffery.schroeder@faa.gov. Associate Fellow.

‡Simulation Engineer, Simulation Laboratories, NASA Ames Research Center, Moffett Field, CA, 94035; william.w.chung@nasa.gov. Senior Member.

II. Previous Transfer of Training Research

Airlines accomplish the vast majority of their training today in simulators with motion cueing in all axes. Specifically, of the roughly 800 training devices that the FAA qualifies, approximately 80% have six-degree-of-freedom motion systems.⁷ That reality stems primarily from three factors: 1) the presence of motion cues adds to pilot acceptance; 2) research shows a subjective preference for motion, even while lacking the concomitant objective measures, and 3) the myriad of cue combinations leads regulators to adopt a conservative approach that tries to make simulators as veridical as possible.

Quests to achieve more cost-effective training solutions, as well as community unease of requiring motion without an objective basis, have led to a variety of transfer-of-training studies or reviews of the literature.^{2,8,9} True transfer-of-training studies divide pilots into groups that train in a simulator with a particular motion condition that may or may not include motion. Then, the training groups fly the same tasks in a real aircraft to determine if any differences arise among the groups.

To date, previous true transfer-of-training studies used general aviation^{10,11} or military aircraft,^{12,13} but no study used a commercial transport aircraft. Such studies are expensive, require many pilots to establish sufficient experimental power, disrupt training curricula, can have safety risks, and are fraught with many experimental pitfalls. Instead, quasi-transfer studies offer an alternative. These studies use a simulator with motion, as opposed to a real aircraft, for evaluating the pilot groups. If a quasi-transfer study shows a relevant difference between pilots that train with motion versus pilots that train without motion, then it is reasonable to assume that training with motion is important. If an experiment does not detect a relevant difference, then it needs to have tested enough pilots to be confident in concluding that having motion makes no difference.

Discussed below are several relevant quasi-transfer-of-training experiments for commercial transport applications. All branches of the military have studied and debated the potential value of motion; however, brevity precludes their discussion. At a high level, many of the military findings are consistent with the studies below.

A quasi-transfer experiment in a six-degree-of-freedom Level C hexapod simulator using a twin-turboprop aircraft showed no operationally relevant differences of using motion during training.¹⁴ The 42 crews experienced an engine out on takeoff. This simulator, although FAA qualified, did not provide the lateral acceleration arising from yawing motion, which can be a key cue for this maneuver. However, many of the hexapods that try to provide that cue often do so poorly, owing to their available simulator displacement or motion drive algorithms.

To address this possible limitation, the same researchers conducted a follow-on quasi-transfer experiment in a six-degree-of-freedom Level D 747-400 simulator.¹⁵ Forty 747-400 pilots flew four tasks, which were an engine out at V_1 , an engine out at V_2 , a precision approach with an engine out and shifting crosswinds, and a sidestep landing with a vertical gust. The researchers adjusted the motion system to maximize the motion fidelity in all axes for the tasks. The study found several statistical differences, although some unexpected. When evaluated in the simulator with motion after their training, the pilots who trained with motion a) responded more slowly on the pedals and had less column motion in the V_2 cut; b) had higher roll and yaw root mean square (RMS) values, and tracked localizer, heading, and airspeed more poorly in the precision approach; c) had increased wheel magnitudes, and lower wheel and pedal control bandwidths in the precision approach; d) had higher yaw RMS values, poorer localizer tracking, more wheel reversals, more wheel deflection and less pedal bandwidth in the sidestep landing; and e) touched down softer and longer in the landing maneuvers. Thus, in many instances, pilots training without motion appeared to learn how to extract more information from the non-motion cues than the pilots who trained with motion. Then, upon adding motion, that non-motion learning provided an incremental benefit. This suggests that, for at least these four tasks, training without motion can provide a performance benefit, and that the average pilot seamlessly uses motion to his benefit when it becomes available. Whether this result is generalizable to an aircraft for these, or all, tasks is unknown.

The previous simulation study used pilots who had flown the actual airplane. To evaluate possible training effects of motion on pilots who had not flown the airplane, another quasi-transfer experiment divided 49 new-hire pilots into a motion and no-motion training group in a Level D 717-200 simulator.¹ The pilots flew two tasks: an engine failure at V_1 , and an engine-out precision approach with shifting winds. The results showed that, during the checking, pilots who trained with motion used less column and had better control of airspeed than the pilots who trained without motion. During the precision approach, the pilots trained with motion used larger pedal inputs. The study did not detect other differences between the motion and no-motion training conditions. Subjectively, pilots rated training with motion more acceptable than those who trained without motion.

The studies discussed above all used task performance in aircraft flight tasks to determine the value of motion for training. In a transfer-of-training study using an abstract compensatory roll tracking task, researchers applied a pilot modeling approach to investigate the effects of simulator motion delays on manual control behavior and tracking performance during various stages of training.¹⁶ Different groups trained without motion or motion with time delays

of 0, 80, 200, and 300 ms compared to the visual cues. After training, all groups transferred to a condition with synchronous visual and motion cues. The results showed that motion affects the acquisition of manual control skills, and that increases in human operator gain, decreases in high-frequency phase lag, and decreases in operator remnant accompanies that learning. Furthermore, tracking performance after the transfer was statistically better for the 80 and 200 ms conditions, relative to the no-motion and 300 ms conditions.

If broad summary statements are possible for the experiments utilizing aircraft flight tasks, then the general situation is: 1) pilot subjective ratings reveal a preference for training with motion; 2) no true transfer-of-training experiment exists with a commercial airplane; 3) in quasi-transfer experiments, the motion available in typical training simulators does not affect the majority of the objective performance measures in aircraft flight tasks; and 4) for the minority of flight task performance measures that are influenced, more often than not, training with no motion results in better performance when evaluated with motion. However, in transfer-of-training experiments that focus solely on the development of manual control skills in target tracking tasks, training with higher fidelity motion enhances learning and improves tracking performance after transfer.

The current experiment improves on the quasi-transfer experiments utilizing aircraft flight tasks. While it does not use a commercial airplane as the evaluation platform, it uses a simulator with an order-of-magnitude more motion than a typical training simulator. The intent is to see whether similar trends found previously in the task performance measures still hold. In addition, the experiment evaluates the effectiveness of recently proposed objective motion criteria and uses those criteria to select two of the four motion conditions.

III. Simulation Setup

III.A. Tasks

The current study used four tasks: an approach and landing with a sidestep, a high-altitude stall recovery, an overbank upset recovery, and an engine-out on takeoff. The takeoff and landing tasks were similar to maneuvers used in a previous transfer-of-training study.¹⁵ The two other tasks were recoveries from unusual attitudes and developed using the Airplane Upset Recovery Training Aid.¹⁷ The design assumption was that motion would help pilots perform these relatively challenging tasks. Fig. 1 shows the flight cards for the four tasks and the following sections describe each in detail.

III.A.1. Approach and Landing with Sidestep

This task began at an altitude of 1,250 ft on a -3 deg glideslope approach to RWY 28R at San Francisco International Airport (SFO). After breaking out of the cloud ceiling at 1,100 ft, air traffic control instructed the pilots to sidestep to SFO RWY 28L. At a distance of 1.6 mi from RWY 28L, the experimenters inserted an upward gust field of 25 ft/sec. Pilots had to maintain the -3 deg glideslope through the sidestep maneuver and the upward gust. After the decision altitude of 200 ft, the task stipulated that the pilots land within a 750-ft-long box with a sink rate of 6 ft/s or less. An audio call-out began at a main gear height of fifty feet and repeated in decrements of 10 ft until touchdown.

This task evaluated if motion affects 1) lateral-directional control, 2) speed and flightpath control along the -3 deg glideslope through the upward vertical gust, and 3) the final approach and touchdown performance.

III.A.2. High Altitude Stall Recovery

The task started with the aircraft cruising at 210 kts ($M = 0.75$) at an altitude of 41,000 ft. The air density at high altitude significantly reduces the engines efficiency and its effectiveness in a stall recovery. Instructions were for the pilots to initiate a self-induced stall by setting the throttles to idle, rolling 15 deg left and pitching approximately 15 deg nose up, decelerating through the stall warning until a negative sink rate was developed (as the aircraft model did not have a pronounced pitch break). To recover from the stall, pilots had to follow the correct recovery sequence of reducing the angle-of-attack (by pitching to approximately 15 deg nose down), leveling the wings, and applying full throttle until establishing a positive climb rate. The task called for the pilots to pull the nose up gently and smoothly so as not to activate the stall warning during the recovery.

This task evaluated if motion affects the recovery performance by helping a pilot damp the flight path response, as well as stabilize the progressively less-stable roll dynamics near stall.

(a) Approach and landing with sidestep.

Task: Approach and landing to SFO RWY 28R with sidestep to visual landing on RWY 28L

Initial Condition: 1,250 ft altitude at $V_{ref} + 5$ on the GS and LOC for RWY 28R

Configuration: gear down, flaps 30 deg

Ceiling/visibility: 1,100 ft ceiling, 10-mile visibility

Wind: 230/12

Turbulence: none

Gusts: vertical

Procedure:

1. Track the GS and LOC to SFO RWY 28R maintaining 141 KIAS
2. Perform sidestep to RWY 28L at ATC command
3. Continue visual to RWY 28L maintaining GS
4. Flare and touchdown 750 –1,500 ft from the threshold
5. Task evaluation ends at touchdown

Desired performance:

1. Deviation from 141 KIAS within ± 5 kts until 200 ft altitude
2. Glideslope deviation within \pm one dot until 200 ft altitude
3. Longitudinal touchdown 750-1,500 ft from threshold
4. Lateral touchdown ± 14 ft from centerline
5. Sink rate at touchdown ≤ 6 ft/s

(b) High altitude stall recovery.

Task: Recover from a high altitude stall

Initial Condition: level at FL 410 altitude and 210 KIAS

Configuration: clean

Ceiling/visibility: CAVU

Wind: none

Turbulence: none

Gusts: none

Procedure:

1. Retard throttle to idle
2. Roll left to a 15 deg bank angle
3. Pull up to decelerate at approximately 4 kt/s
4. Continue deceleration through stick shaker until a sink rate develops
5. Apply nose down pitch, roll as needed, power as needed to return to steady-state flight

Desired performance:

1. Maintain bank angle within ± 5 degs about 15 during entry
2. Proper stall recovery procedure (push, roll, thrust, stabilized flight)
3. Recovers to speed > 210 KIAS and positive rate of climb
4. Recovers without exceed airplane's limitations
5. Oscillatory load factor peaks between 0.75 and 1.25 g
6. Recovers with smooth and positive control
7. During recovery, no more than one additional stick shaker activation

(c) Overbank upset recovery.

Task: Recover from 120 deg bank and 20 deg nose down pitch attitude

Initial Condition: level at 5,000 ft altitude and 200 KIAS

Configuration: clean

Ceiling/visibility: CAVU

Wind: none

Turbulence: none

Gusts: none

Procedure:

1. Hands off controls as computer flies the aircraft to 120 deg left bank and 20 deg nose down pitch attitude
2. At the command "your airplane", unload, roll wings level, and then apply nose-up pitch rate similar to a takeoff rotation rate to recover safely to level flight

Desired performance:

1. Pilot unloads before rolling
2. Pilot does not initiate pull up before approaching wings level
3. Pilot does not exceed 2.5 g in the pull out
4. Pilot adjusts thrust to idle if needed to prevent overspeed (< 350 kts)

(d) Engine out on takeoff.

Task: Recover from engine out on takeoff

Initial Condition: at takeoff position on RWY 28R

Configuration: gear down, flaps 10 deg

Ceiling/visibility: CAVU

Wind: none

Turbulence: none

Gusts: none

Procedure:

1. Advance throttles to takeoff thrust ($60\% N_1$)
2. Maintain centerline
3. Rotate at $V_r = 128$ kts to a pitch attitude of 10 degs and establish speed of $V_2 + 10$
4. Maintain heading and speed after single engine failure

Desired performance:

1. Desired heading ± 5 degs
2. Desired airspeed ± 5 kts
3. Establishes a bank of approximately 5 degs towards operating power-plant

Figure 1. Experiment flight cards.

III.A.3. Overbank Upset Recovery

The task started at a speed of 200 kts and an altitude of 5,000 ft. The simulator automatically flew the aircraft into the initial upset condition with a bank angle of 120 deg and 20 deg nose down. After reaching this condition, an audio message indicated that pilots were now in control. The task required that the pilots follow the correct recovery sequence by unloading (pushing column forward), leveling the wings, and then applying a nose-up pitch rate similar to a takeoff rotation rate while staying within the aircraft's 2.5 g operational limit. All of this needed to be accomplished without impacting the ground.

This task evaluated if motion helped the recovery performance primarily through modulating of load factor.

III.A.4. Engine Out on Takeoff

Upon a takeoff roll, the task asked for the pilots to rotate at 128 kts, establish a 10 deg nose up pitch attitude, and maintain a climb speed of 145 kts (V_2+10). Either the left or right engine failed randomly when the main landing gear left the ground. After identifying the failed engine, the pilots needed to apply near full rudder pedal input towards the good engine, roll approximately 5 deg into the direction of the failed engine to maintain the desired heading, and modulate the remaining thrust to maintain speed.

This task evaluated if motion could help the pilots detect the failed engine and promptly and correctly recover using appropriate lateral-directional control.

III.B. Aircraft Model

The experimenters modified an existing mid-size twin-engine commercial transport aircraft model with a gross weight of 185,800 lbs for this study. The enhancements allowed for a more representative aircraft response in the stall and overbank tasks. Most significantly, modifications to the lift coefficient as a function of angle of attack allowed for typical transport post-stall characteristics at high angles of attack. Modifications to the rolling moment coefficient due to roll rate allowed for satisfactory representation of reduced roll instability near stall.

The simulated aircraft had six degrees of freedom and a two-crew cockpit with conventional flight controls. The aircraft had conventional control surfaces such as ailerons, spoilers, flaps, rudders, and elevators. The aircraft model included a full landing gear model, allowing it to taxi on the ground, and takeoff and land. The aircraft could operate in the full flight envelope up to a maximum cruising altitude of 43,000 ft.

III.C. Motion Cueing

The experiment had a total of seven different motion conditions: a no-motion condition, two hexapod motion conditions, and a separately optimized VMS motion condition for each of the four tasks. The VMS motion conditions utilized as much of the VMS motion space as possible for each task, creating a so-called truth motion case, which likely represents the best that a ground-based simulator can do. The hexapod motion conditions simulated a six-degree-of-freedom hexapod motion platform, similar to ones used on current Level-D training simulators. The high-fidelity large hexapod motion condition used as much motion space of the simulated hexapod as possible. The low-fidelity small hexapod motion condition was a copy of the high-fidelity hexapod condition, but with motion gains reduced by 50%. The low-fidelity hexapod motion condition was basically representing a motion system half the size of the high-fidelity hexapod motion system. Table 1 summarizes the seven motion conditions. This table also contains the labels for each motion condition used throughout the rest of the paper.

Table 1. Motion conditions.

Label	Description
NOM	no-motion for all tasks
HLO	small hexapod motion for all tasks
HHI	large hexapod motion for all tasks
SIDE	VMS motion for sidestep task
HALT	VMS motion for stall task
OVER	VMS motion for overbank task
TOFF	VMS motion for takeoff task

III.C.1. Motion Logic and Tuning

Fig. 2 depicts the overall motion logic. The standard VMS motion algorithm controlled the VMS motion conditions. Some details are provided in Ref. 18. The hexapod motion conditions utilized a motion algorithm developed by Parrish et al.¹⁹ However, deliberate disabling of the adaptive part of the algorithm turned the adaptive filters into constant washout filters for the translational accelerations and the rotational rates. The algorithms calculated the leg extensions of the hexapod, thus accounting for the inherent constraints when trying to simultaneously move the hexapod in several degrees of freedom. Since the VMS has independent degrees of freedom, the resulting position and orientation commands after accounting for a possible leg reaching its limit had to be calculated. A Newton-Raphson

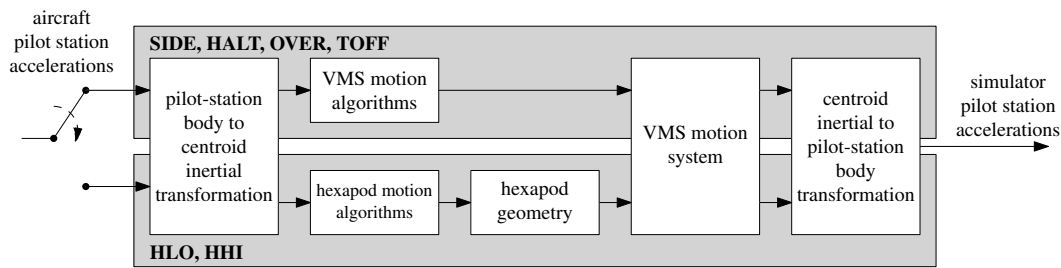


Figure 2. Objective motion cueing test setup.

optimization took the leg extensions and solved for the platform position and orientation in real-time. Next, the motion platform positions, rates and accelerations were sent to the VMS motion system (Fig. 2). The maximum extension of the hexapod legs was 60 inches. Mass and inertia effects of the hexapod system were not taken into account. As such, identical hardware dynamics resulted for each motion condition. The equivalent time delays of the VMS motion system for the pitch, roll, yaw, longitudinal, lateral, and vertical axes are, 47, 68, 48, 50, 69, and 67 ms, respectively.¹⁸

Again, the washout filter settings for the low- and high-fidelity hexapod motion conditions were the same for every task, except for the 50% reduction in gains when going from HHI to HLO. The Sinacori criteria served as the guide to tune the hexapod motion parameters in combination with a simulation model of the hexapod motion logic.^{20–22} That process resulted in washout filter settings as close to the medium- or high-fidelity regions of the Sinacori criteria as possible without reaching motion limits by using a representative run of each flight task. Some subjective tuning followed the initial selection of the hexapod parameters.

The Sinacori criteria also served as the guide for tuning the VMS washout filters.^{20,21} The optimization process consisted of applying the criteria while using as much of the VMS motion space as possible. Prepositioning the simulator cab to one side or the other of its displacement or orientation limits allowed for additional fidelity improvements. For example, for the engine out on takeoff, prepositioning the cab laterally in the direction to the opposite side of the failed engine allowed for additional available lateral travel.

Fig. 3 depicts Sinacori criteria for all the motion conditions and all degrees of freedom. The fidelity boundaries in the diagrams are from Ref. 21. The VMS motion conditions generally had higher fidelity than the hexapod conditions. Note that the sway, heave, and yaw degrees of freedom show the biggest difference in fidelity between the hexapod and VMS motion conditions. The big differences for the sway and heave degrees of freedom are a direct result of the large lateral and vertical travel of the VMS. The differences in other degrees of freedom are mainly due to the uncoupled degrees of freedom in the VMS motion system, as opposed to coupled ones in a hexapod system. That is, the same set of legs produce all the motions, so when at a particular pitch orientation, less leg travel is available for roll.

III.C.2. Objective Motion Cueing Test

The objective motion cueing test (OMCT) is a standardized test to evaluate the performance of a simulator motion system.⁴ In recent years, various organizations applied the test to different motion-base simulators at simulator manufacturers, training facilities, and research institutes.^{23,24} Based on these preliminary test data, a group proposed a set of motion criteria.³ The OMCT considers the translational accelerations and rotational rates at the pilot station. The results of the OMCT are the frequency responses relating the translational accelerations and rotational rates at the aircraft pilot station to those at the simulator pilot station, the inputs and outputs of the block diagram in Fig. 2. These responses provide a full description of the dynamic behavior of the motion system in the frequency domain.

The authors performed the OMCT for all motion conditions in this experiment. Results and fidelity regions based on the proposed motion criteria are given in Fig. 4. For brevity, the figure omits the test results for the cross coupling between lateral and rotational axes. The plots in Fig. 4 also depict the regions encompassed by one standard deviation on either side of the mean OMCT measurements of eight different research and training simulators. Again, the hexapod (HLO and HHI) and VMS motion conditions (SIDE, HALT, OVER, and TOFF) use different motion logic, as depicted in Fig. 2. However, all motion conditions utilized the same hardware.

Fig. 4 shows that the OMCT frequency responses of the hexapod motion conditions are mostly in, or near, the regions encompassed by one standard deviation on either side of the mean of the preliminary OMCT data. However, larger deviations occurred for the pitch, surge, and sway axes. This is mainly caused by the inherent coupling between the surge and pitch, and sway and roll filters as a consequence of the tilt coordination implementation in the hexapod

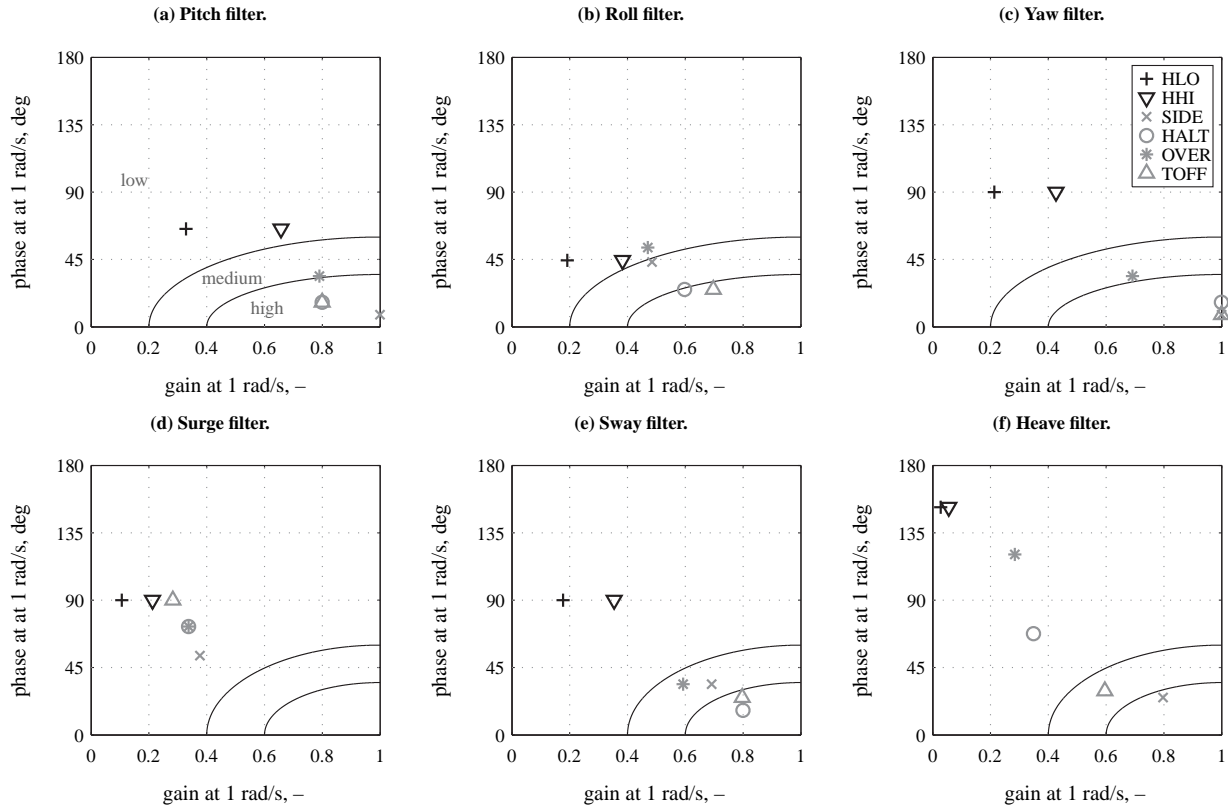


Figure 3. Sinacori diagrams for all motion conditions.

logic.¹⁹ The VMS motion conditions had significantly higher gains and smaller phase differences across the whole frequency range for all axes. The heave and yaw degrees of freedom show the biggest difference in fidelity between the hexapod and VMS motion conditions. Compared to the Sinacori criteria (Fig. 3), the difference in fidelity for the sway degree of freedom appears to be less significant. This is mainly caused by the contribution to the frequency response of the tilt coordination at lower frequencies.

IV. Experiment Design

IV.A. Method

IV.A.1. Independent Variable

The experiment had one independent variable: the motion training condition. There were four different training motion groups: no motion, low-fidelity hexapod, high-fidelity hexapod, or VMS. After training, each pilot repeated the task in a check run with the VMS truth motion case. Thus, the experiment was a between-subjects design with every pilot being assigned to one of the four training groups. Table 2 provides the training and check motion conditions used for each group. Section III.C discusses the motion conditions in more detail.

Table 2. Experiment motion groups.

Motion Group	Sidestep Task		Stall Task		Overbank Task		Takeoff Task	
	Training	Check	Training	Check	Training	Check	Training	Check
1. No Motion	NOM	SIDE	NOM	HALT	NOM	OVER	NOM	TOFF
2. Low-Fidelity Hexapod	HLO	SIDE	HLO	HALT	HLO	OVER	HLO	TOFF
3. High-Fidelity Hexapod	HHI	SIDE	HHI	HALT	HHI	OVER	HHI	TOFF
4. VMS	SIDE	SIDE	HALT	HALT	OVER	OVER	TOFF	TOFF

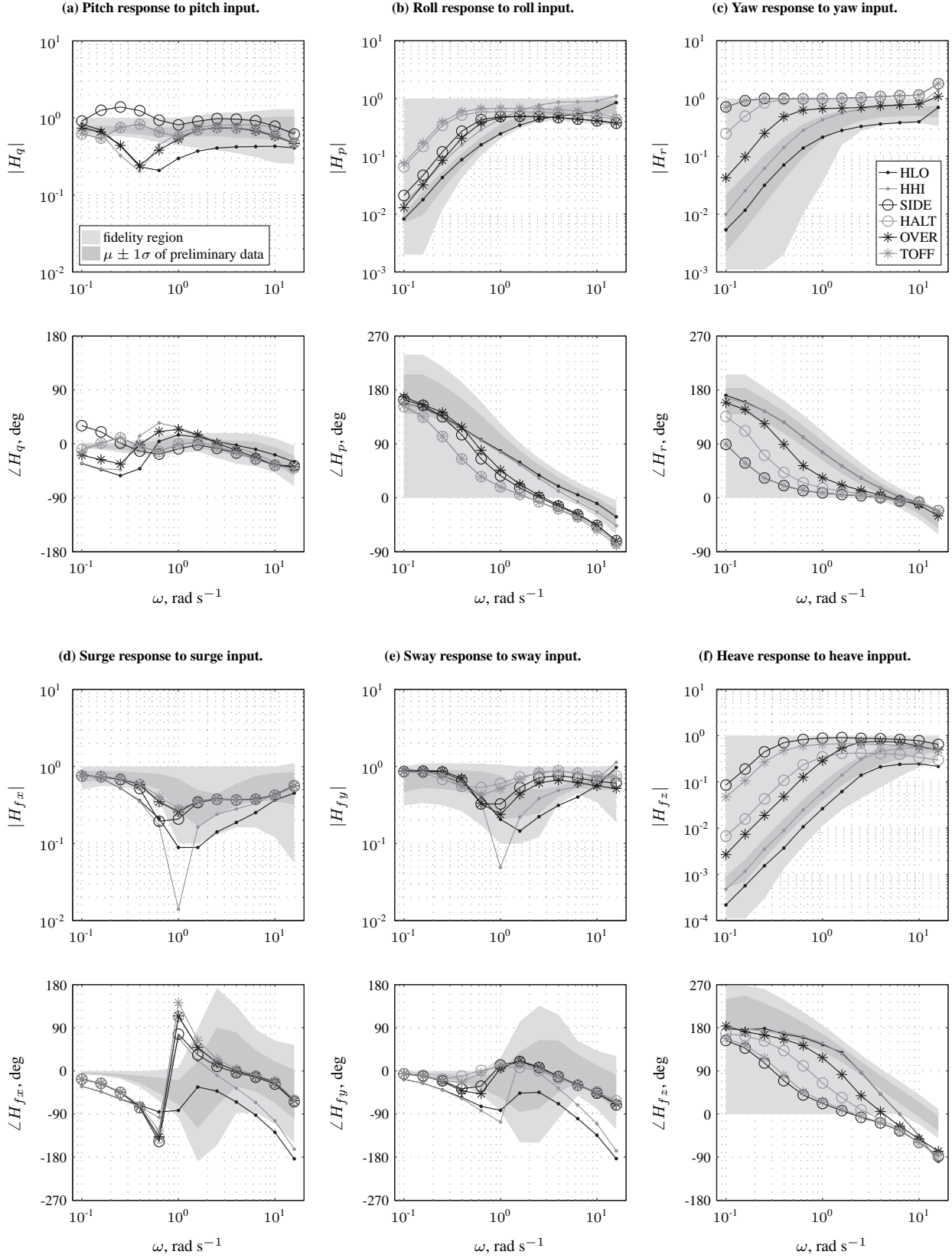


Figure 4. Objective motion cueing test results for all motion conditions.

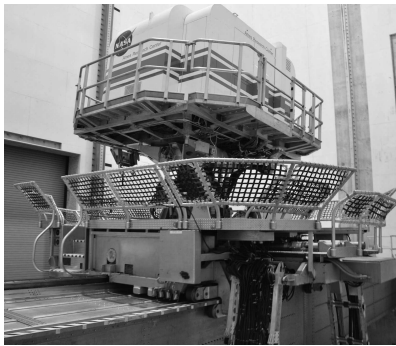


Figure 5. Vertical Motion Simulator.



Figure 6. Cockpit setup.

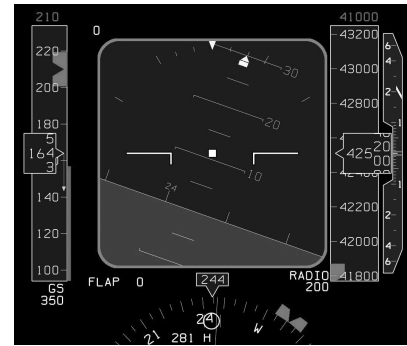


Figure 7. Primary flight display.

IV.A.2. Apparatus

The experiment used the VMS (Fig. 5). The capabilities of the motion system can be found in Ref. 18. The experiment used the transport cab (T-CAB). Pilots flew from the left seat; the instructor occupied the right seat. Both pilots had three displays in front of them. Positioned to the left was a navigation display, a primary flight display (PFD) in the middle, and an engine display to the right (Fig. 6). In the center of the cab, an additional display showed task performance after each run. The instructor pilot used this screen to provide performance feedback to the participant pilot. This screen was blank when the participant pilot was flying.

The control column and wheel controlled aircraft pitch and roll, respectively. A thumb switch on the wheel controlled elevator trim. Conventional rudder pedals controlled the rudder deflection in the air and nose wheel rotation on the ground. The forces on each control approximated those of a medium-sized airliner. The throttle quadrant had four throttle levers. However, pilots used only the left-most two, closest to them, as the simulated aircraft had only two engines. The pilot could extend the flaps and gear with representative controls for those functions. However, the instructor pilot configured these controls prior to the start of each task.

The PFD symbology was similar to the symbology on a Boeing 777 PFD (Fig. 7). Speed and heading bugs indicated the desired speed and heading for each task. In addition, typical symbology on the speed tape indicated the minimum and maximum speeds, and V-speeds. The PFD also depicted conventional localizer and glideslope error indicators. A green speed trend vector originating from the current indicated airspeed showed what the airspeed would be in 10 s. The control columns in the cockpit were equipped with stick shakers to warn pilots for an impending stall. The activation of the shakers occurred simultaneously when the minimum speed tape, also known as the barber pole, coincided with the current indicated airspeed.

The collimated out-the-window display of T-CAB consisted of six cathode ray tube (CRT) monitors casting a high-quality image on six spherical mirrors. The mirrors formed a dome-like section providing a continuous field-of-view image to both pilots. The out-the-window visual had a 220° horizontal field of view and a 30° vertical field of view (10° up and 20° down). A Rockwell-Collins EPX5000 computer image generator created the out-the-window visual scene. The visual system equivalent time delay was 62 ms.¹⁸ This was in line with the equivalent time delays of the motion system (Section III.C).

IV.A.3. Procedures

Before the start of the experiment, pilots received an extensive briefing on the purpose and overall procedures. In addition, the briefing explained the flight conditions, procedures, and performance criteria for each flight task. The briefing informed the pilots that the training runs would use a particular motion condition and the check runs would always use the highest fidelity motion condition (Table 2). They were told that the motion in the training versus the check could be different or the same. They were not given any specifics about the training motion condition.

Pilots performed all four tasks. Each pilot performed the training for each task with only one of the four motion conditions. After the training for one task, pilots performed a check run for the same task with the VMS motion condition. A randomized Latin square determined the order of the tasks. The number of training runs for each task was dependent on the participant pilot reaching proficiency, as determined by the performance criteria set for each task (Fig. 1). Pilots flew a minimum of three training runs and a maximum of six. Their training was complete once they reached proficiency. However, if a pilot reached proficiency before three tries, they continued until they had flown the

task at least three times. Owing to the challenges of the tasks, pilots could fly the tasks a maximum of six times prior to being given the check.

Before each task, the instructor pilot briefed the participant pilot in the VMS cab about the upcoming task. This briefing reviewed the procedures and performance criteria of the task. In addition, the instructor explained all the relevant controls and display symbology. The instructor evaluated a pilot's performance after each run. An experiment observer in the control room verified the evaluation. After completing each check run, participant pilots rated the motion with a rating scale. Section IV.A.5 gives more information on the rating scale.

IV.A.4. Participants

Sixty-one general aviation pilots participated. As such, this group of pilots represents those who would be receiving initial, instead of recurrent, training. Thirty-five pilots had a private pilot license (PPL) and 26 a commercial pilot license (CPL). Forty pilots had an instrument rating. None of the pilots had experience with flying commercial transport aircraft. The experimenters assigned the pilots to each motion group semi-randomly. The initial selection for a motion group was random. However, for the final selection, we took into account a pilot's total number of flying hours to create groups that had similar flying hour averages. Three pilots had more than 2,000 flying hours, with one pilot having more than 10,000 hours. Not taking into account these three pilots, the average number of flying hours was 554 for the no-motion group, 664 for the low-fidelity hexapod group, 470 for the high-fidelity hexapod group, and 450 for the VMS motion group. There was a large variation between pilots in the number of simulator hours. However, only one or two of the pilots had flown a motion-base simulator.

IV.A.5. Dependent Measures

Four subjective dependent measures and 13 objective dependent measures were recorded and analyzed. Pilots rated the simulator motion in the check run compared to the motion in the training run. Every pilot gave four of these subjective ratings in total, one for each task. Pilots had to rate the following statement: "The motion cues helped me perform this task better during the check than they did during the training." Their response could either be: 1) strongly disagree, 2) disagree, 3) somewhat disagree, 4) neutral, 5) somewhat agree, 6) agree, or 7) strongly agree.

Several objective outcome-based variables determined the effect of training motion on task performance in the check run. Many measures related directly to the performance criteria for each task (Fig. 1). For the approach and landing with sidestep, the RMS of the glideslope, RMS_{gs} , and speed deviation in the approach, RMS_V , were calculated. Calculations for these variables used data from the last 60 s before reaching the decision altitude of 200 ft. When the main gear touched the runway, data captures occurred for the sink rate, \dot{h}_{td} , and the longitudinal and lateral deviations, Δx_{td} and Δy_{td} , from the desired touchdown point.

For the stall recovery, the RMS roll error, RMS_ϕ , applied to the 30 s before the start of the stall recovery. The maximum pitch rate deviation, q_{max} , and the number of stick shakers, N_s , applied to the stall recovery segment of the run. Pitch rate was substituted as a partial surrogate for load factor, as it was more suitable for analyzing oscillatory behavior in the stall recovery. Maximum load factor and maximum airspeed during the recovery, n_{z-max} and V_{max} , were performance measures for the overbank upset recovery task. These variables came from data in the recovery segment. Finally, for the engine out on takeoff task, the RMS of the heading and speed deviations after the engine failure, RMS_ψ and RMS_V , used data from the last 15 s after the engine failure. The reaction time of the initial pedal input after the engine failure, t_p , was from the time of the engine failure to the time when the pedal input was 10% of the maximum input.

IV.B. Hypotheses

The VMS can provide an order of magnitude more physical motion than the simulators used in previous quasi-transfer-of-training studies, so we expected to find more significant differences between motion groups in the task performance data of the check run.

In transfer-of-training studies solely focusing on the development of manual control skills, training with higher fidelity motion did significantly improve performance after transfer.¹⁶ For this reason, we hypothesized that pilots training with VMS motion would perform better overall in the check run than pilots training without motion or with hexapod motion. This is partly due to the fact that motion enhances learning of manual flying skills,¹⁶ but also because pilots having trained without motion or hexapod motion might be startled when first experiencing the large motion excursions in the VMS motion condition during the check run.

More specifically, we hypothesized that pilots would learn to address the vertical gust in the approach task and the engine failure in the takeoff task more effectively when training with higher fidelity motion, leading to lower RMS values of the glideslope or heading deviations in the check run (RMS_{gs} and RMS_{ψ}), as motion cues have been shown to have an alerting effect when rapid changes in attitude rates occur.²⁵ For the same reason, we also expected the pedal reaction time, t_p , after the engine failure to be lower in the check run when pilots trained with higher fidelity motion. Furthermore, we hypothesized that pilots training with higher fidelity motion would learn to address the sink rate in the landing flare, \dot{h}_{td} , and the load factor in the overbank recovery maneuver, n_{z-max} , more effectively. We expected pilots training with higher fidelity motion to more effectively compensate for disturbances when keeping the roll angle at 15 deg in the approach to the stall, allowing for lower RMS_{ϕ} values during the check run.¹⁶ Finally, we hypothesized that pilots training with higher fidelity motion would trigger less additional stick shakers, N_s , in the stall recovery during the check run, as it would allow them to more effectively manage pitch rate and load factor oscillations. We expected the remaining dependent measures to be less affected by training motion.

V. Results

The results presented are from the check-run data only; that is, for all the data presented, it was the last run of each task that all pilots flew with the VMS motion condition. The motion-condition labels in the following figures indicate the motion condition used for training (Table 2). Error-bar plots present the continuous interval data with means and 95% confidence intervals. Bar plots present ordinal data showing the median of each group. Gray dashed lines depict the task performance criteria, wherever appropriate.

The analyses used one-way ANOVAs to detect statistically significant differences in continuous interval data and Kruskal-Wallis tests to detect significant differences in ordinal data. For a one-way ANOVA to be able to provide a valid result, three assumptions must hold: 1) the data have no outliers in any groups, 2) data in each group are normally distributed, and 3) groups must have equal variances.

To comply with the first assumption, inspection of the box plots for each task identified data outliers. Any data points that were more than 1.5 box-lengths away from the edge of their box were classified as outliers and removed from subsequent data analyses. The paper does not present the box plots. The analysis employed the Shapiro-Wilk test of normality to determine if the data were normally distributed across groups. The outlier removal caused most of the resulting data to be normally distributed. However, in a few cases, the data needed to be deskewed with a data transformation.

Finally, Levene's Test of Homogeneity of Variance evaluated the assumption of homogeneity of variances. If variances were homogeneous across groups, a normal one-way ANOVA tested for significant differences between motion groups, and a Tukey post-hoc test was performed for multiple comparisons between groups. If the assumption of homogeneity of variances was violated, the analysis used a more robust Welch one-way ANOVA, in combination with a Games-Howell post-hoc test for multiple comparisons. All statistical tests had a significance level of 0.05. Table 3 provides a summary of the statistical test results for all dependent measures of the experiment. The remainder of this section will provide a more detailed overview of the results for each task.

Table 3. Summary of statistical test results. KW indicates Kruskal-Wallis test, AN indicates one-way ANOVA, and WA indicates one-way Welch ANOVA. ** indicates a significant effect ($p < 0.05$), * indicates a marginally-significant effect ($0.05 \leq p < 0.10$), and – is not significant ($p \geq 0.10$).

Measure	Test	df	F/χ^2	p	Sig.
Sidestep Task					
rating	KW	3.0	2.210	0.530	–
RMS_{gs}	AN	3.0,51.0	1.430	0.245	–
RMS_V	AN	3.0,51.0	2.152	0.105	–
\dot{h}_{td}	AN	3.0,51.0	0.865	0.465	–
Δx_{td}	AN	3.0,51.0	3.400	0.025	**
Δy_{td}	WA	3.0,27.2	2.959	0.050	*
Stall Task					
rating	KW	3.0	1.585	0.663	–
RMS_{ϕ}	AN	3.0,53.0	0.971	0.413	–
q_{max}	AN	3.0,53.0	1.020	0.391	–
N_s	KW	3.0	11.151	0.011	**
Overbank Task					
rating	KW	3.0	7.893	0.048	**
n_{z-max}	AN	3.0,53.0	2.687	0.056	*
V_{max}	AN	3.0,53.0	1.452	0.238	–
Takeoff Task					
rating	KW	3.0	12.199	0.007	**
RMS_{ψ}	WA	3.0,30.1	0.629	0.602	–
RMS_V	AN	3.0,56.0	0.948	0.424	–
t_p	AN	3.0,56.0	3.142	0.032	**

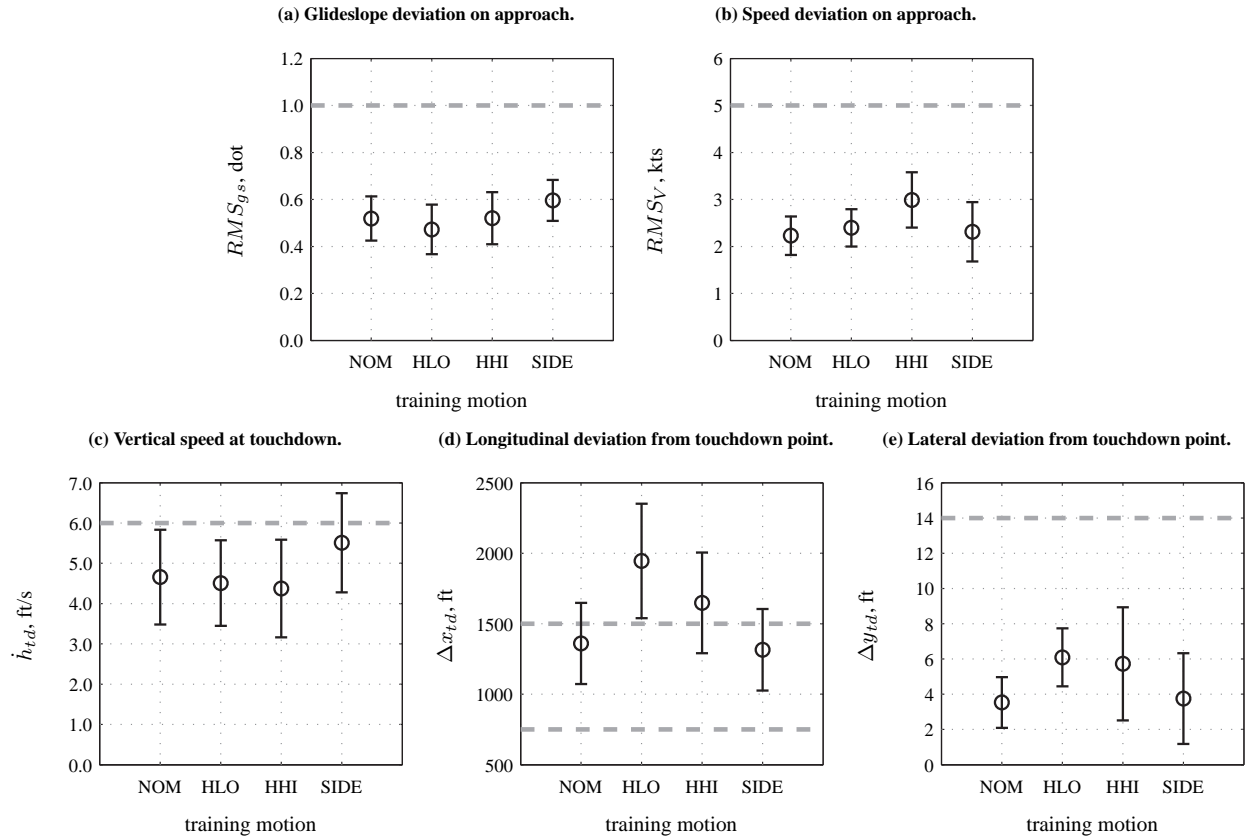


Figure 9. Performance data of the approach and landing with sidestep task.

V.A. Approach and Landing with Sidestep

The above process identified six pilots as outliers, and the results do not include their data for the approach and landing task. This left 14 pilots in the no-motion and hexapod motion groups and 13 pilots in the VMS motion group. Only eight pilots trained to proficiency on this task.

Fig. 8 depicts the motion cueing ratings given by pilots after the check run (Section IV.A.5). The bars indicate the number of times pilots gave a certain rating. The plot shows a large variation in ratings, especially for the groups who trained with no-motion (NOM) and low-fidelity hexapod motion (HLO). A Kruskal-Wallis test did not detect statistically significant differences among the different groups (Table 3).

Fig. 9 depicts the pilot performance data for the approach and landing task. Figs. 9a and 9b show the RMS of the glideslope deviation and the speed deviation during the approach, respectively. Figs. 9c, 9d, and 9e depict the sink rate at touchdown, and the longitudinal and lateral distances from the desired touchdown point, respectively. Glideslope deviation RMS, speed deviation RMS, and the longitudinal and lateral deviation from the desired touchdown point were positively skewed. Taking the square root of the data deskewed these variables. The data transformation and the removal of the 6 outliers resulted in a normal distribution for all motion groups. Furthermore, there was homogeneity of variances between groups for all variables, except for the lateral deviation from the touchdown point.

There were no statistically significant differences between motion groups for the glideslope deviation RMS, speed deviation RMS, and the sink rate at touchdown. The longitudinal deviation from the desired touchdown point was statistically significantly different between motion groups. Tukey post-hoc analysis revealed that groups who trained with the HLO and SIDE motion conditions were significantly different ($p = 0.038$), and the groups who trained with NOM and HLO were marginally-significantly different ($p = 0.052$). The lateral deviation from the desired touchdown

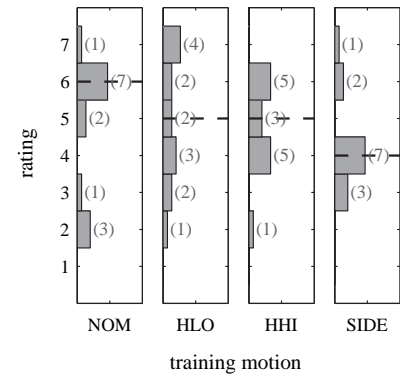


Figure 8. Pilot ratings for the approach and landing with sidestep task.

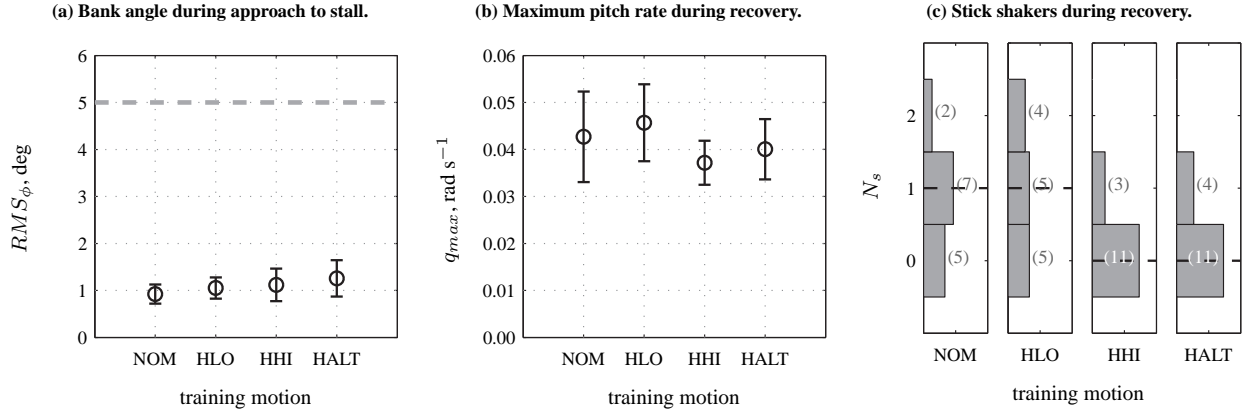


Figure 11. Performance data of the high altitude stall recovery task.

point is also statistically significantly different between motion groups. Games-Howell post-hoc analysis revealed that only the difference between the groups training with NOM and HLO was marginally significant ($p = 0.061$). Comparing Figs. 9d and 9e reveals a similar effect in longitudinal and lateral deviation from the desired touchdown point; that is, an increased lateral deviation is visible with increases in the longitudinal deviation.

V.B. High Altitude Stall Recovery

The outlier removal process found four outlier pilots. The high altitude stall task data analysis does not include their data. This left 14 pilots in the groups training with NOM, HLO and HHI motion and 15 in the group training with HALT motion. Of these 57 pilots, 56 trained to proficiency on this task.

Fig. 10 shows the motion ratings given by pilots. The ratings have a large spread in all motion groups. However, in all groups, pilots seem to be neutral or agree with the statement that motion cues during the check run helped them to perform the task better. The Kruskal-Wallis test did not detect a statistically significant difference in the ratings between the motion groups.

Fig. 11 shows the performance data of the high altitude stall recovery task. The roll deviation RMS during the stall entry, maximum pitch rate deviation during the recovery, and the number of stick shaker activations during the recovery are given in Figs. 11a, 11b, and 11c, respectively. A square root transformation deskewed the pitch rate data. Both the roll deviation RMS and the transformed maximum pitch rate were normally distributed, and homogeneity of variances existed between groups for both variables. The roll deviation RMS during the stall entry was not significantly different between motion groups. The maximum pitch rate deviation in the recovery was also not significantly different between motion groups.

Fig. 11c depicts the number of stick shakers triggered during the stall recovery for each motion group. The bars indicate how many pilots triggered a certain amount of shakers during the recovery. The bar plot shows that more pilots triggered additional stick shakers when training with the NOM and HLO motion conditions compared to the HHI and HALT conditions. This effect is highly significant, with $\tilde{N}_s = 1$ for the groups training with NOM and HLO motion and $\tilde{N}_s = 0$ for the groups training with HHI and HALT. Pairwise comparisons revealed that only the number of stick shakers in groups who trained with the HLO and HHI motion conditions were statistically significantly different. Although no significant differences between groups are detected for q_{max} , a similar trend might be observed in q_{max} and N_s .

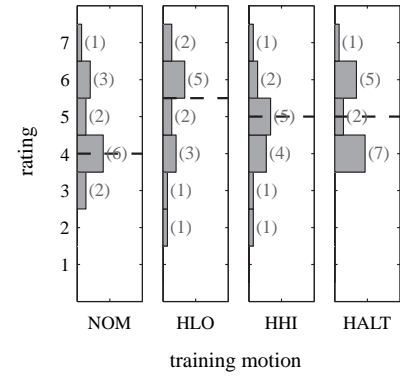


Figure 10. Pilot ratings for the high altitude stall recovery task.

V.C. Overbank Upset Recovery

For the overbank upset recovery task, four pilots were outliers and removed from the data analysis. After removal of these outliers, the VMS motion group contained 15 pilots, and the remaining motion groups 14. Of these pilots, a total of 29 trained to proficiency on this task.

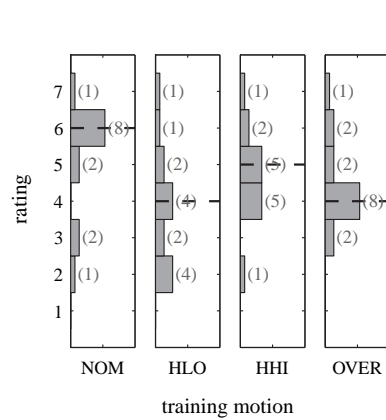


Figure 12. Pilot ratings for the overbank task.

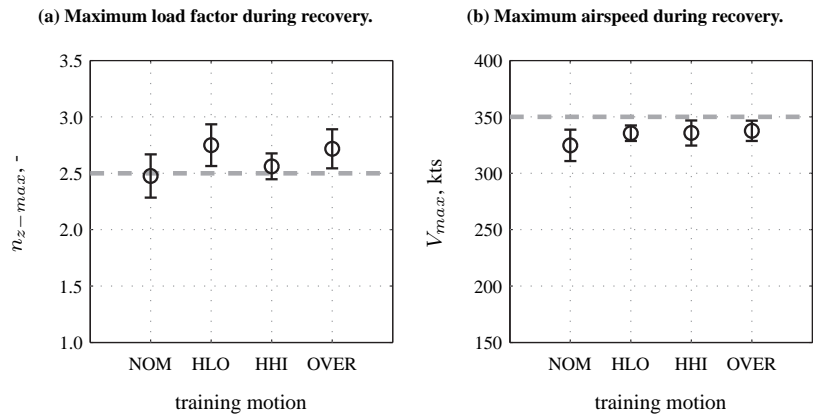


Figure 13. Performance data of the overbank upset recovery task.

Fig. 12 provides a bar plot of the motion ratings. The ratings are spread out for every motion group, but more so for the groups who trained with NOM and HLO motion. The analysis detected a statistically significant difference between motion groups. The ratings were more neutral for the groups training with motion, meaning that pilots in these groups thought that the usefulness of the motion in performing the task was comparable in the training or the check. Pairwise comparisons revealed that only the groups training with NOM and HLO motion were statistically significantly different compared to each other ($p = 0.037$).

Fig. 13 gives the performance measurements for the overbank upset recovery task. The maximum load factor and the maximum airspeed during the recovery are given in Figs. 13a and 13b, respectively. The maximum load factor and the maximum airspeed had normal distributions and homogeneity of variances between groups. The difference in maximum load factor obtained was marginally significant between motion groups. Post-hoc analysis revealed that this significant effect is mainly between the groups training with NOM and HLO motion ($p = 0.083$). For the groups that trained with motion, the maximum load factor was higher in the check run compared to the group that trained without motion. The maximum airspeed was not significantly different between motion groups. However, similar to the maximum load factor, an increase in maximum airspeed between the no-motion group and motion groups might be observed.

V.D. Engine Out on Takeoff

One pilot was an outlier and removed from the data analysis for the engine out on takeoff task. This left 14 pilots in the no-motion, 15 pilots in low-fidelity hexapod and VMS motion groups, and 16 pilots in the high-fidelity hexapod motion group. On this task, 35 pilots trained to proficiency.

Fig. 14 displays the motion ratings for the engine out on takeoff task. Again, a large spread in ratings exists for each motion group. However, in the groups who trained with NOM and HLO, pilots agreed that the motion cues in the check run helped them perform the task better as compared to the motion cues in the training runs. For the groups training with HHI and TOFF the answer was more neutral. This difference between groups was statistically significant. A pairwise comparison revealed that the group training with NOM is statistically significantly different from the groups training with HHI and TOFF ($p = 0.007$ and $p = 0.036$, respectively).

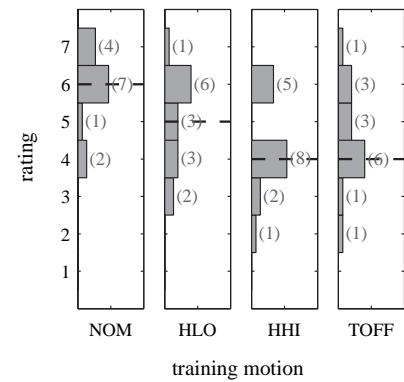


Figure 14. Pilot ratings for the engine out on takeoff task.

Fig. 15 shows the performance data for the engine out on takeoff task. Fig. 15a depicts the RMS of the deviation from the desired heading, Fig. 15b the RMS of the airspeed deviation, and Fig. 15c the pedal reaction time. Taking the square root of the speed and heading deviation data deskewed it. After the data transformation, both variables had a normal distribution for every group, except for the heading deviation in the VMS motion group. The speed deviation had homogeneity of variances between groups, but not the heading deviation. The pedal input reaction time had a normal distribution and homogeneous variances between all groups.

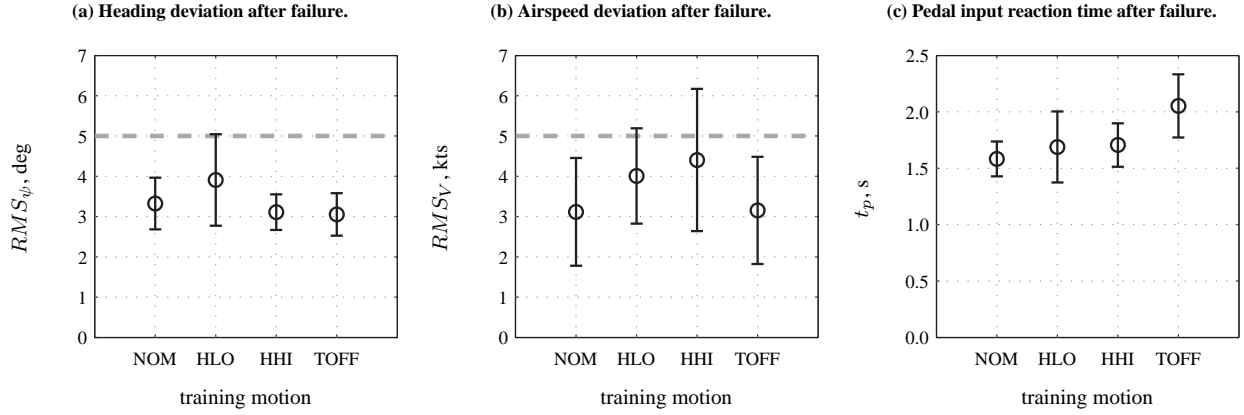


Figure 15. Performance data of the engine out on takeoff task.

The analysis did not detect a significant difference in heading deviation RMS after engine failure between motion groups. There were also no significant differences detected between motion groups for the speed deviation RMS after the engine failure. However, there were significant differences in pedal input reaction time between motion groups. The pedal input reaction time increased from the group training with NOM ($\bar{t}_p = 1.58$ s), HLO ($\bar{t}_p = 1.69$ s), HHI ($\bar{t}_p = 1.71$ s), to TOFF ($\bar{t}_p = 2.05$ s). This overall significant effect is mainly caused by the significant difference between the groups training with NOM and TOFF motion ($p = 0.03$), as revealed by the post-hoc analysis.

VI. Discussion

VI.A. Effects of Motion on Transfer of Training

We chose to only present data from the check run for each task to focus on how motion in training affects task performance after transfer. Pilots performed a single check run. Thus, we were not able to average check runs, a common way to reduce the variability in measurements. Furthermore, it was difficult for pilots to meet all the performance criteria in most of the tasks, especially as none had experience in transport category aircraft. We did not allow pilots to get familiar with the aircraft dynamics before the start of the task. Despite this, we still found some significant differences.

Each task had a large spread in motion ratings. This is a common problem with subjective ratings, as they are dependent on personal preferences and experiences. Pilots had to rate if the motion cues during the check helped them fly the task better than what they experienced during training (Section IV.A.5), a difficult task in itself, as it required pilots to remember the motion they experienced in previous runs. Some pilots indicated that the question was unclear to them, which could have led to rating variability. However, the overbank and takeoff tasks still yielded significant differences in the ratings. In both tasks, pilots in the no-motion training group agreed that the motion in the check helped them to perform the task better, while the answer was closer to neutral for the other groups. A plausible reason for these differences is that, during the check, the additional vertical cues in the overbank and the additional lateral cues in the engine-out takeoff aided in moderating load factor and sideslip, respectively. The high-altitude stall recovery task had the least consistent rating results. The motion for this task was less intense compared to the motion during some parts of other tasks.

In the sidestep landing task, the training motion condition affected the longitudinal and lateral touchdown point in the check with VMS motion. However, the trends in these variables, although similar in both, were not consistent with the increase in motion fidelity. Pilots who trained with small hexapod motion landed outside the desired touchdown zone, while pilots training with VMS motion landed in the zone. Pilots who trained with small hexapod motion strayed from the centerline more than those who trained without motion, although both groups met the performance criterion. Fig. 9d shows that pilots generally touched down long. This is most likely the result of pilots not being familiar with landing an aircraft the size of an airliner. Only eight of the 55 pilots trained to proficiency on this task. This might have contributed to the lack of significant differences detected.

In the high-altitude stall recovery, the training motion condition affected how many times a pilot reactivated the stall warning when checked with VMS motion. Pilots training without motion, or with the small hexapod motion condition, obtained another stall warning in the recovery during the check, while pilots who trained with large hexapod

or VMS motion did not. This significant difference is most likely caused by the much higher pitch motion fidelity in the HHI and HALT motion conditions. With higher-fidelity pitch motion, pilots may have been able to use those cues to provide lead compensation in pitch, thereby resulting in less angle-of-attack oscillations during the recovery, and in turn less stick shaker activations. The lack of significant differences between groups for the roll deviation RMS in the approach to the stall was unexpected, given the results of a previous transfer-of-training experiment utilizing a roll tracking task.¹⁶ This might be caused by the fact that the bank angle performance criterion was not strict enough.

In the overbanked upset recovery, the training motion condition marginally affected the maximum load factor obtained when checked with VMS motion. Pilots who trained without motion pulled less g in the check than those who trained with small hexapod motion. It appears that pilots training without motion learned to utilize the visual information more effectively and thus had an advantage when presented with full motion in the check.

Finally, for the engine out on takeoff, the training motion condition affected pedal reaction time in the check. During the check, pilots who trained without motion applied pedal inputs faster than the pilots who trained with VMS motion. This result was also found in a previous transfer-of-training study.¹⁵ Again, a possible explanation for this might be that pilots training with lower fidelity motion learned to use the visual cues more effectively, and they used this advantage in the check run.

In summary, we found significant effects of training motion fidelity on objective measures in the check for three of the four tasks and a marginally significant effect on the fourth task. For the stall recovery task, pilots training with higher fidelity motion appeared to have an advantage after transfer. Interestingly, this is also the task where almost every pilot trained to proficiency. However, for the overbank recovery and engine out on takeoff, pilots training with higher fidelity motion appeared to perform worse after transfer. Only around half of the pilots trained to proficiency for these tasks. These latter results are different from what we hypothesized based on the effects of motion on skill acquisition (Section IV.B) and more in line with the results of a previous study.¹⁵ Some aspects of the experiment design likely contributed to only finding 5 significant, and 2 marginally significant, motion effects in the 17 measures.

One aspect was the scripted nature of each task, allowing pilots to anticipate critical events. For example, some pilots started to fly low just before the start of the vertical gust in the approach and landing task, making it more easy to meet the performance criterion for the glideslope deviation. Another aspect might be that pilots had no experience with the aircraft dynamics, making it difficult to meet the performance criteria in the already challenging tasks. Finally, the performance criteria, or our instructions about the criteria, might not have been strict enough. For example, the instructions “keep the bank angle at 15 deg at all times” and “keep the bank angle between 10 and 20 deg” induce different behaviors and, consequently, effects of motion. However, most of these aspects of the experiment design are dictated by how pilot training in flight simulators is currently performed.

We frequently observed pilots training with no motion or hexapod motion being surprised by the large motion cueing in the check, causing them to perform significantly worse compared to the training runs. We performed analyses of variance with many of the pilot demographics we collected (such as, total hours and the type of rating) as covariates, to correct for confounding variables. However, none of the covariates significantly affected the results. There might have been additional confounding variables that we were not able to account for (for example, we didn’t collect pilots’ age).

Finally, task performance might not be the best measure for transfer of training, as humans are highly adaptive and able to achieve similar performance under different conditions. Measures that are able to capture this human adaptation, such as pilot model parameters, might be more suitable.¹⁶ However, these measures often induce many constraints on the design of the task used for training, making the task less similar to real piloting tasks.

VI.B. Objective Motion Cueing Test and Criteria

This study was the first to test the effectiveness of the newly proposed objective motion cueing criteria in assessing simulator motion fidelity. The OMCT allowed for calculation of the frequency responses relating aircraft and simulator pilot station motions for all conditions (Section III.C.2). The OMCT specification describes sine waves at 12 different frequencies to be used as input test signals.⁴ For the specific force and rotational inputs, different sine amplitudes are provided for every frequency. However, for the majority of the frequencies, the amplitude for the specific force input is 1 m/s^2 and for the rotational input 1 deg/s . Due to the large number of motion conditions we needed to test, we used multi-sine signals instead of single sine waves, allowing us to test all frequencies at the same time for a given degree of freedom and motion condition. Optimization of the individual sine input amplitudes allowed for measurements with satisfactory signal-to-noise ratios at all frequencies without exceeding simulator excursion limits.

The OMCT specification provides standardized signals and procedures to perform the test. However, this one-size-fits-all approach might not be appropriate in every case. The maneuvers of our experiment produced large aircraft motions and, as a result, the motion filter gains had to be reduced considerably, especially for the hexapod motion

filters. In our case, the amplitudes provided in the OMCT specification were not adequate to get measurements with sufficient signal-to-noise ratios in all axes. Some amplitudes were significantly increased to get accurate results. In addition, the OMCT requires motion filters to be linear and constant over time. However, many flight simulators used for training utilize adaptive motion filters with time-varying motion gains. This indicates that good knowledge of the motion and measurement systems is required for accurate results. It might be useful to introduce some self checks in the OMCT specification to allow for a validation of the results. For example, by repeating the test with different amplitudes, more insight could be gained into nonlinear effects and signal-to-noise ratios.

The OMCT results provide a full description of the basic dynamic behavior of the motion system over a wide frequency range, as opposed to the Sinacori criteria, which only provide data at 1 rad/s. In addition to the OMCT, we also used the Sinacori criteria to compare the different motion conditions.^{20,21} Comparing the two methods reveals that the OMCT results sometimes provide a less dramatic difference between motion conditions with a large variation in fidelity, especially for the axes where rotational and translational channels both affect the frequency response. For example, the Sinacori criteria reveal a large difference in fidelity in the sway degree of freedom between the hexapod and VMS motion conditions (Fig. 3e). However, this difference appears to be less significant in the OMCT results (Fig. 4e).

The recently proposed objective motion criteria (the fidelity regions in Fig. 4) are based on the OMCT measurements of 8 different research and training simulators. For some axes, our OMCT measurements are located outside of the fidelity regions. This is mainly caused by the inherent structure of the hexapod and VMS motion logics we used; that is, it was impossible to find motion filter parameters that result in a frequency response within the criterion boundaries. This makes the criteria less suitable for the tuning of more unconventional motion filters and might warrant specific criteria for different types of motion logic.

VII. Conclusions

This experiment used the NASA Ames Vertical Motion Simulator (VMS) with different motion conditions to evaluate whether or not training with motion is valuable for initial training of commercial transport pilots. The study also provided the first evaluation of newly proposed objective motion criteria. Sixty-one general aviation pilots flew four challenging tasks. After training each pilot on each task with one of four motion conditions, the pilots repeated each task in a check utilizing the large motion envelope of the simulator.

For the subjective measures, the detected trends were conventional. That is, 1) pilots who trained without motion during an overbank upset recovery believed that VMS motion helped them more than pilots who trained with small-hexapod motion; and 2) pilots who trained without motion in the engine-out takeoff believed that VMS motion helped them more than pilots who trained with large-hexapod or VMS motion.

For the objective measures, training with more motion fidelity helped during the check in some tasks and hurt on others. Specifically, 1) the average pilot who trained with VMS motion landed in the touchdown zone, while the average pilot who trained with small hexapod motion landed outside of it; 2) pilots who trained with large-hexapod motion had fewer stall warnings when recovering from a high-altitude stall than pilots who trained with small-hexapod motion.

Having better motion fidelity during training hurt when checked in the following cases: 1) the average pilot who trained without motion landed in the touchdown zone, while the average pilot who trained with small hexapod motion landed outside of it; 2) the average pilot who trained without motion landed closer to the centerline than pilots who trained with small hexapod motion; 3) the average pilot who trained without motion during an overbank upset recovery stayed within the g-limit, while the average pilot who trained with small-hexapod motion exceeded it; 4) pilots who trained without motion had faster pedal reaction times during the VMS check than pilots who trained with VMS motion.

In the instances where lower motion fidelity during training resulted in improvements during the check, one hypothesis is that pilots may have learned control strategies in relatively poorer cueing situations and then carried those strategies over to the check when the motion fidelity improved. Whether such learned strategies have unintended consequences is a question that needs addressed but was beyond scope.

The two principal conclusions were: 1) the study found objective measures that depended on the motion condition in training; and 2) the new objective motion criteria show early promise, as the predicted fidelity improvement between the two conditions covered by the criteria, the small and large hexapod conditions, resulted in expected trends in pilot ratings and objective performance measures after transfer.

Acknowledgments

The authors thank the simulation engineers at NASA Ames SimLabs who contributed to the experiment. We especially thank Michael Weinstein, Deanna Doan, Emily Lewis, Nick Riccobono, Bosco Dias, Scott Reardon, and Boris Rabin for their valuable contributions. We also thank Gordon Hardy for being a test pilot and an instructor pilot. Finally, we appreciate the continued program support of Dr. Andrew Cheng at the FAA Technical Center.

References

- ¹Bürki-Cohen, J. and Go, T. H., "The Effect of Simulator Motion Cues on Initial Training of Airline Pilots," *Proceedings of the AIAA Modeling and Simulation Technologies Conference and Exhibit, San Francisco (CA)*, No. AIAA-2005-6109, 15–18 Aug. 2005.
- ²de Winter, J. C. F., Dodou, D., and Mulder, M., "Training effectiveness of whole body flight simulator motion: A comprehensive meta-analysis," *The International Journal of Aviation Psychology*, Vol. 22, No. 2, April 2012, pp. 164–183.
- ³Hosman, R. J. A. W. and Advani, S. K., "Are Criteria for Motion Cueing and Time Delays Possible? Part 2," *Proceedings of the AIAA Modeling and Simulation Technologies Conference, Boston (MA)*, No. AIAA-2013-4833, 19–22 Aug. 2013.
- ⁴International Civil Aviation Organization, ICAO 9625: *Manual of Criteria for the Qualification of Flight Simulation Training Devices. Volume 1 – Aeroplanes*, 2009, 3rd edition.
- ⁵Code of Federal Regulations, *Flight Simulation Training Device Initial and Continuing Qualification and Use*, Title 14, Part 60.
- ⁶National Transportation Safety Board, Washington, D.C., *In-Flight Separation of Vertical Stabilizer, American Airlines Flight 587, Airbus Industrie A300-605R, N14053, Belle Harbor, New York, November 12, 2001*, Oct. 2004, Aircraft Accident Report NTSB/AAR-04/04.
- ⁷Federal Aviation Administration, *FAA Qualified Simulators by Make and Model*, May 2014, http://www.faa.gov/about/initiatives/nsp/train_devices/media/All_FSTD.pdf.
- ⁸Bürki-Cohen, J., Sparko, A. L., and Bellman, M., "Flight Simulator Motion Literature Pertinent to Airline-Pilot Recurrent Training and Evaluation," *Proceedings of the AIAA Modeling and Simulation Technologies Conference, Portland (OR)*, No. AIAA-2011-6320, 08–11 Aug. 2011.
- ⁹McCauley, M. E., "Do Army Helicopter Training Simulators Need Motion Bases?" Technical report 1176, Naval Postgraduate School, Feb. 2006.
- ¹⁰Koonce, J. M., "Effects of Ground-Based Aircraft Simulator Motion Conditions Upon Prediction of Pilot Proficiency. Part I," Tech. Rep. AFOSR-TR-74-1292, University of Illinois, April 1974.
- ¹¹Jacobs, R. S., "Simulator Cockpit Motion and the Transfer of Initial Flight Training," Tech. Rep. ARL-76-8/AFOSR- 76-4, University of Illinois, June 1976.
- ¹²Martin, E. L. and Waag, W. L., "Contributions of Platform Motion to Simulator Training Effectiveness: Study I Basic Contact," Tech. Rep. AFHRL-TR-78-15, Brooks Air Force Base, June 1978.
- ¹³Pohlmann, L. D. and Reed, J. C., "Air-to-Air Combat Skills: Contribution of Platform Motion to Initial Training," Tech. Rep. AFHRL-TR-78-53, Brooks Air Force Base, Oct. 1978.
- ¹⁴Go, T., Bürki-Cohen, J., and Soja, N. N., "The Effect of Simulator Motion on Pilot Training and Evaluation," *Proceedings of the AIAA Modeling and Simulation Technologies Conference and Exhibit, Denver (CO)*, No. AIAA-2000-4296, 14–17 Aug. 2000.
- ¹⁵Bürki-Cohen, J. S., Go, T. H., Chung, W. W., and Schroeder, J. A., "Simulator Platform Motion Requirements for Recurrent Airline Pilot Training and Evaluation," Tech. rep., Human Factors Research and Engineering Division, John A. Volpe National Transportation Center, Sept. 2004.
- ¹⁶Levison, W. H., Lancraft, R. E., and Junker, A. M., "Effects of Simulator Delays on Performance and Learning in a Roll-Axis Tracking Task," *Fifteenth Annual Conference on Manual Control*, Wright State University, Dayton (OH), 20–22 March 1979, pp. 168–186.
- ¹⁷Federal Aviation Administration, *Airplane Upset Recovery Training Aid Revision 2*, Nov. 2008.
- ¹⁸Beard, S. D., Reardon, S. E., Tobias, E. L., and Aponso, B. L., "Simulation System Optimization for Rotorcraft Research on the Vertical Motion Simulator," *Proceedings of the AIAA Modeling and Simulation Technologies Conference, Minneapolis (MN)*, No. AIAA-2012-4634, 13–16 Aug. 2012.
- ¹⁹Parrish, R. V., Dieudonne, J. E., Bowles, R. L., and Martin, Jr., D. J., "Coordinated Adaptive Washout for Motion Simulators," *Journal of Aircraft*, Vol. 12, No. 1, 1975, pp. 44–50.
- ²⁰Sinacori, J. B., "The Determination of Some Requirements for a Helicopter Research Simulation Facility," Tech. Rep. NASA CR-152066, Systems Technology, Inc., Sept. 1977.
- ²¹Schroeder, J. A., "Helicopter Flight Simulation Motion Platform Requirements," Tech. Rep. NASA/TP-1999-208766, NASA, July 1999.
- ²²Gouverneur, B., Mulder, J. A., van Paassen, M. M., Stroosma, O., and Field, E. J., "Optimisation of the SIMONA Research Simulator's Motion Filter Settings for Handling Qualities Experiments," *Proceedings of the AIAA Modeling and Simulation Technologies Conference and Exhibit, Austin (TX)*, No. AIAA-2003-5679, 11–14 Aug. 2003.
- ²³Roza, M., Meiland, R., and Field, J., "Experiences and Perspectives in Using OMCT for Testing and Optimizing Motion Drive Algorithms," *Proceedings of the AIAA Modeling and Simulation Technologies Conference, Boston (MA)*, No. AIAA-2013-4835, 19–22 Aug. 2013.
- ²⁴Stroosma, O., van Paassen, M. M., Mulder, M., Hosman, R. J. A. W., and Advani, S. K., "Applying the Objective Motion Cueing Test to a Classical Washout Algorithm," *Proceedings of the AIAA Modeling and Simulation Technologies Conference, Boston (MA)*, No. AIAA-2013-4834, 19–22 Aug. 2013.
- ²⁵Young, L. R., "Some Effects of Motion Cues on Manual Tracking," *Second Annual NASA-University Conference on Manual Control*, Massachusetts Institute of Technology, Cambridge (MA), 28 Feb.–2 March 1966, pp. 231–240.